THE EMBO JOURNAL

Manuscript EMBO-2010-76582

Integrative epigenomic mapping defines four major chromatin states in Arabidopsis

François Roudier, Ikhlak Ahmed, Caroline Bérard, Alexis Sarazin, Tristan Mary-Huard, Sandra Cortijo, Daniel Bouyer, Erwann Caillieux, Evelyne Duvernois-Berthet, Liza Al-Shikhley, Laurène Giraut, Barbara Després, Stéphanie Drevensek, Frédy Barneche, Sandra Dèrozier, Véronique Brunaud, Sébastien Aubourg, Arp Schnittger, Chris Bowler, Marie-Laure Martin-Magniette, Stéphane Robin, Michel Caboche and Vincent Colot

Corresponding author: Vincent Colot, IBENS

Review timeline:

Submission date:	23 November 2010
Editorial Decision:	21 December 2010
Revision received:	09 March 2011
Editorial Decision:	10 March 2011
Accepted:	10 March 2011

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 21 December 2010

Thank you for submitting your manuscript for consideration by The EMBO Journal. It has been now been evaluated by three referees and I enclose their reports below. As you will see the referees find the analysis of the organization of the Arabidopsis genome to be interesting and important and recommend publication in The EMBO Journal once several issues are clarified, this includes a more direct comparison with previous studies including the recent study from the van Steensel lab. Given the interest from the referees should you be able to address these issues, we would be happy to consider a revised manuscript.

I should remind you that it is EMBO Journal policy to allow a single round of revision only and that, therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript. When you submit a revised version to the EMBO Journal, please make sure you upload a letter of response to the referees' comments. Please note that when preparing your letter of response to the referees' comments that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website: http://www.nature.com/emboj/about/process.html

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #1 (Remarks to the Author):

This manuscript reports the systematic genome-wide mapping of a broad set of 11 histone marks plus DNA methylation in Arabidopsis seedlings. This unique dataset enabled the authors to conduct a detailed integrative analysis to understand the fine patterns and relationships among these marks. Among others, the authors find identify major combinations of marks (or actually three; the fourth lacks essentially all tested marks).

The strength of this manuscript lies in the thorough and systematic approach, which to my knowledge has not yet been done in any plant species. The data and all analyses are very robust. While the manuscript does not report many surprises, these kind of systematic surveys are very important for the "big picture". I therefore recommend (in principle) that it be published in EMBO Journal.

Major points:

- The data are restricted to histone marks (and 5mC), while none of hundreds of chromatin-associated proteins are considered here. I therefore feel that the claim to have identified "major chromatin types" (title, abstract) is premature. Twelve marks is also not enough to be sure that the entire 'space' of chromatin types has been sampled. It therefore would be more appropriate/accurate to use the term "major combinations of histone marks".

- The authors chose "well-characterized antibodies". Please substantiate this statement by adding a supplementary table that lists for each antibody how its specificity was confirmed in previous studies (peptide blots? peptide competition? etc), including relevant citations. Furthermore, it seems that the authors don't fully trust the specificity of the H3K27me2 and H3K9me3 antibodies. While I very much appreciate the authors' honesty about these doubts, I recommend that the most obvious candidates for crossreactivity (H3K27me1 and H3K27me3 for the H3K27me2 antibody, and H3K36me3 for the H3K9me3 antibody) are directly tested by peptide spotblot.

Minor points:

- It seems to me that CT4 may correspond to "black" chromatin as identified by Filion et al. Black chromatin also appears to lack known histone marks (so far).

- Fig1C: ordering along the horizontal axis seems arbitrary. Clustering is better visualized if both axes are ordered the same way.

- Fig4D: add grey scale bar for expression level

Referee #2 (Remarks to the Author):

The authors profiled eight histone modifications (H3K4me2 and 3, H3K27me1 and 2, H3K36me3, H3K56ac, H4K20me1, and H2Bub) of Arabidopsis seedlings, and took a systematic analysis both genome-wise and gene-wise, in combination with four other epigenetic maps (H3K9me2 and 3,

H3K27me3 and DNA methylation) obtained before. The authors identified the 12 epigenetic marks as 2 distinct groups by pair-wise association, distributing mainly in euchromatic regions or heterochromatic regions. They also described the biased association of some of the epigenetic marks with genes in terms of the transcription activity and gene length, as well as their spatial distribution relative to the gene annotations. They further identified 4 major chromatin types by c-means clustering based on the associations between the marks, which represent the heterchromatic regions, the actively transcribed chromatins, the repressed chromatins by PcG proteins, and ambiguous chromatins. Apart from the extended domains of heterchromatic regions, the euchromatic arms are largely interspersed by the 4 major types of chromatins. Finally, the authors discussed the chromatin composition in relation to expression specificity, focusing on the apparently double-indexed genes by both H3K4me3 and H3K27me3.

The data available to the community is valuable, and it's interesting to see the distinct types of chromatin compositions.

Major points:

- The value of the data sets as a ressource for the community is limited because most data are based on a tiling array that covers only one of the 5 chromosomes and has a relatively low resolution (~1kb). Having said that, it should be noted that the major conclusions are justified.

- Results part 1, 2 and 4 should be condensed. Given the over-interpretation often found with apparent bivalent chromatin in plants, the corresponding section of the manuscript is refreshing, however it does not fit well with the other parts. A much shortened version of this section would fit better into the discussion.

- The authors mentioned sub-types existing beneath the 4 major types. It would be very interesting to see the more detailed categorization.

- The section on gene length effects remained unclear to the reviewer and requires major re-writing. Examples of statements that confused the reviewer are "In particular,whereas H2Bub, H3K36me3 and H3K27me1 show similar distribution over the transcribed region of genes independently of their length, H2Bub and H3K36me3 are poorly detected over genes smaller than ~1.5 kb." and "Cytosine methylation is not frequently found over small genes either, although H3K27me3 deposition is somewhat biased towards smaller genes"

Minor points:

- In Figure 3A, how is Expression percentiles defined? Are all gene uniformly distributed in each percentile? If so then why would the "All genes" dashed line show an non-uniform distribution?

- On page 8, paragraph 2: "two of these four types change little over k values..." which two of the types change and which other don't? Explain clearly or do not mention it.

- On page 12, paragraph 2: "H3K79me3...a chromatin mark that is apparently missing in (Zhang et al, 2007)". Missing in Arabidopsis?

- "about 10% of the tiles do not show association with any of the chromatin modifications" - Failure to detect signals can always be a technical issue. The section describing and discussing missing signals should be shortened and phrased more carefully or even removed.

Referee #3 (Remarks to the Author):

The manuscript of Roudier et al. describes the genome wide view of 10 different chromatin marks, including selected histone marks and DNA methylation marks, and thus provides an organizational map of the Arabidopsis epigenome. Moreover, the genome wide maps of eight of these marks were generated for the first time in this study and these data will provide a very important resource for the epigenetic community, especially for plant epigeneticists. The authors discovered that the Arabidopsis epigenome can be confined into four major types of chromatin, which are not only

clearly associated with specific combinations of epigenetic marks but can also be well distinguished as different functional elements.

This is a very important and clearly presented study that provides readers with an essential and novel message.

There is only one minor point that should be addressed prior publication: the authors should provide a thorough discussion of their presented results and conclusions in the context of the similar, previously published studies performed for other organisms. This is especially important for the very recent discovery of five chromatin domains in Drosophila (Filion et al Cell 143 , 2010). Although the study of Filion at al,. and also others, are mentioned in the discussion the significance of the similarities and differences in the results presented in these papers with the results of the present work needs to be systematically discussed and interpreted. Such discussion including comparison of experimental approaches, number and sizes chromatin domains, their chromosomal distribution and functional relationships across various organisms would make this paper much more attractive for readers. Obviously this requires serious rewriting of the manuscript, with mentioning of the other studies already in the introduction, and a comparison of results throughout the manuscript: however, this extra effort would pay off in that it would allow the reader to gain a more global view of chromatin organization across different species.

1st Revision - authors' response 09 March 2011

Point by point response:

Referee #1:

Major points:

- The data are restricted to histone marks (and 5mC), while none of hundreds of chromatin-associated proteins are considered here. I therefore feel that the claim to have identified "major chromatin types" (title, abstract) is premature. Twelve marks is also not enough to be sure that the entire 'space' of chromatin types has been sampled. It therefore would be more appropriate/accurate to use the term "major combinations of histone marks".

We would like to emphasize that this study aimed at identifying prevalent chromatin states based on combinations of 12 chromatin marks. Although the 4 main chromatin states identified cover ~95% of the genome ‘space’, we agree with Referee #1 that our dataset cannot ascertain that the entire set of chromatin types has been sampled. To avoid confusion with the denomination ”chromatin types” used to describe the five principal combinations of chromatin-associated proteins identified via a much larger integrative analysis in Drosophila (Filion et al, 2010), we now use the denomination “chromatin states” (CS) and we clearly points in the Discussion to the possibility of additional such states in Arabidopsis. Nonetheless, we wish to stress that we have not only identified four major combinations of chromatin marks but have also determined that they have distinct functional properties.

- The authors chose "well-characterized antibodies". Please substantiate this statement by adding a supplementary table that lists for each antibody how its specificity was confirmed in previous studies (peptide blots? peptide competition? etc), including relevant citations. Furthermore, it seems that the authors don't fully trust the specificity of the H3K27me2 and H3K9me3 antibodies. While I very much appreciate the authors' honesty about these doubts, I recommend that the most obvious candidates for crossreactivity (H3K27me1 and H3K27me3 for the H3K27me2 antibody, and H3K36me3 for the H3K9me3 antibody) are directly tested by peptide spotblot.

We have removed this sentence form the text and we now provide peptide competition assays as recommended (Supplementary Figure 5). We also provide in a table (Supplementary Table IX) available information on commercial antibodies. Finally, we address more specifically the issue of antibody specificity in the Discussion.

Minor points:

- It seems to me that CT4 may correspond to "black" chromatin as identified by Filion et al. Black chromatin also appears to lack known histone marks (so far).

We now provide a more detailed comparison with the Filion et al article and mention the similarity between BLACK and what we now call CS4.

- Fig1C: ordering along the horizontal axis seems arbitrary. Clustering is better visualized if both axes are ordered the same way.

We feel that the directionality of association provides important information given the difference in abundancy between the chromatin marks analyzed. We therefore have included clustering information for the two axes and have amended figures legends accordingly.

- Fig4D: add grey scale bar for expression level

We have added a grey scale bar has been.

Referee #2:

The authors profiled eight histone modifications (H3K4me2 and 3, H3K27me1 and 2, H3K36me3, H3K56ac, H4K20me1, and H2Bub) of Arabidopsis seedlings, and took a systematic analysis both genome-wise and gene-wise, in combination with four other epigenetic maps (H3K9me2 and 3, H3K27me3 and DNA methylation) obtained before. The authors identified the 12 epigenetic marks as 2 distinct groups by pair-wise association, distributing mainly in euchromatic regions or heterochromatic regions. They also described the biased association of some of the epigenetic marks with genes in terms of the transcription activity and gene length, as well as their spatial distribution relative to the gene annotations. They further identified 4 major chromatin types by c-means clustering based on the associations between the marks, which represent the heterchromatic regions, the actively transcribed chromatins, the repressed chromatins by PcG proteins, and ambiguous chromatins. Apart from the extended domains of heterchromatic regions, the euchromatic arms are largely interspersed by the 4 major types of chromatins. Finally, the authors discussed the chromatin composition in relation to expression specificity, focusing on the apparently double-indexed genes by both H3K4me3 and H3K27me3.

The data available to the community is valuable, and it's interesting to see the distinct types of chromatin compositions.

Major points:

- The value of the data sets as a ressource for the community is limited because most data are based on a tiling array that covers only one of the 5 chromosomes and has a relatively low resolution (~1kb). Having said that, it should be noted that the major conclusions are justified.

We acknowledge the fact that referee #2 agrees that the major conclusions reached using the chromosome 4 tiling array are justified. In addition, we would like to stress that the genome-wide data that we present are for 7 of the 12 marks, including H3K36me3 and H2ub, which have not been analyzed at the genome scale before. Also, all of these these data will be publicly available at http://epigara.biologie.ens.fr/index.html , together with relevant epigenomic data published by other groups. This should therefore represent an important epigenomic resource for the community.

- Results part 1, 2 and 4 should be condensed. Given the over-interpretation often found with apparent bivalent chromatin in plants, the corresponding section of the manuscript is refreshing, however it does not fit well with the other parts. A much shortened version of this section would fit better into the discussion.

We have reorganized the Results part into 3 sections. Original parts 2 and 4 (which dealt with chromatin indexing of genes) have now been merged into a single section (section 3), which is much

more condensed. Original parts 1 and 3 (which dealt with the individual and combinatorial analyses of the marks across all genomic sequences) are now sections 1 and 2, and are essentially unchanged. This reorganization makes the results section much easier to read.

- The authors mentioned sub-types existing beneath the 4 major types. It would be very interesting to see the more detailed categorization.

Our intention was simply to mention that additional chromatin states likely exist. As splitting of CS1 by increasing the number of k in c-means clustering is neither robust nor biologically interpretable, we have decided to remove mention of possible additional subtypes/refined states from the Results section. The notion that the broad-level organization provided by our work might be further refined is now addressed more thoroughly in the Discussion.

- The section on gene length effects remained unclear to the reviewer and requires major re-writing. Examples of statements that confused the reviewer are "In particular,whereas H2Bub, H3K36me3 and H3K27me1 show similar distribution over the transcribed region of genes independently of their length, H2Bub and H3K36me3 are poorly detected over genes smaller than ~1.5 kb." and "Cytosine methylation is not frequently found over small genes either, although H3K27me3 deposition is somewhat biased towards smaller genes"

This part has been re-written and significantly shortened.

Minor points:

- In Figure 3A, how is Expression percentiles defined? Are all gene uniformly distributed in each percentile? If so then why would the "All genes" dashed line show an non-uniform distribution?

Whole seedling transcriptome data were retrieved from Schmid et al. (2005) and genes were binned into 20 expression percentiles according to their absolute expression values. For each bin, the number of genes marked by a given chromatin modification was scored and represented as a percentage of all the genes marked by this modification. The dashed line represents the distribution of all annotated genes on chromosome 4 across all expression percentiles.
Legend of Figure 4 has been amended and the procedure is now described in Materials and Methods.

- On page 8, paragraph 2: "two of these four types change little over k values..." which two of the types change and which other don't? Explain clearly or do not mention it.

See above.

- On page 12, paragraph 2: "H3K79me3...a chromatin mark that is apparently missing in (Zhang et al, 2007)". Missing in Arabidopsis?

Yes, this has been corrected.

- "about 10% of the tiles do not show association with any of the chromatin modifications" - Failure to detect signals can always be a technical issue. The section describing and discussing missing signals should be shortened and phrased more carefully or even removed.

This part has been removed.

Referee #3 (Remarks to the Author):

The manuscript of Roudier et al. describes the genome wide view of 10 different chromatin marks, including selected histone marks and DNA methylation marks, and thus provides an organizational map of the Arabidopsis epigenome. Moreover, the genome wide maps of eight of these marks were generated for the first time in this study and these data will provide a very important resource for the epigenetic community, especially for plant epigeneticists. The authors discovered that the Arabidopsis epigenome can be confined into four major types of chromatin, which are not only

clearly associated with specific combinations of epigenetic marks but can also be well distinguished as different functional elements.

This is a very important and clearly presented study that provides readers with an essential and novel message.

There is only one minor point that should be addressed prior publication: the authors should provide a thorough discussion of their presented results and conclusions in the context of the similar, previously published studies performed for other organisms. This is especially important for the very recent discovery of five chromatin domains in Drosophila (Filion et al Cell 143 , 2010). Although the study of Filion at al,. and also others, are mentioned in the discussion the significance of the similarities and differences in the results presented in these papers with the results of the present work needs to be systematically discussed and interpreted. Such discussion including comparison of experimental approaches, number and sizes chromatin domains, their chromosomal distribution and functional relationships across various organisms would make this paper much more attractive for readers. Obviously this requires serious rewriting of the manuscript, with mentioning of the other studies already in the introduction, and a comparison of results throughout the manuscript: however, this extra effort would pay off in that it would allow the reader to gain a more global view of chromatin organization across different species.

We agree with referee#3 that the comparison of the chromatin landscape in different organisms is an interesting topic. We now provide such comparisons in the Abstract, Introduction and Discussion, where we point in some detail to similarities and differences in chromatin indexing between Arabidopsis and metazoans, referring notably to the five recent papers of Kharchenko et al, 2010; Riddle et al, 2010; The modENCODE Consortium/Roy et al, 2010; Ernst & Kellis, 2010 and Filion et al. 2010. However, we have not extended these comparisons to the Results section, as we feel as this would have unnecessarily lengthen the manuscript.

2nd Editorial Decision 10 March 2011

I have read through your revised manuscript and I find that you have satisfactorily incorporated all the changes suggested by the referees, including an interesting comparison with the chromatin organisation in other organisms. I am happy to accept the manuscript for publication in The EMBO Journal. I believe it will make a great contibution to journal.

Yours sincerely,

Editor
The EMBO Journal

